A close-up photograph of a dark blue New York City Police Department uniform. The image shows the collar with a silver star, a shoulder patch with the NYPD logo, and a cluster of gold-colored medals on the left side. The text "POLICE DEPARTMENT" and "CITY OF NEW YORK" is visible on the patch.

# Designing a resource plan for the NYPD

Jackson Au

# Crime is prevalent in New York!

New York has steady cases of crime.

This is something the New York Police Department is constantly trying to lower over the years.

Rate of Crime per 100,000 people	
Year	Crime Cases
2014	2,101.8
2015	1,986.8
2016	1.921.8
2017	1.858.7
2018	1.791.0

# PROBLEM STATEMENT

Though 2,000 cases per 100,000 people may seem small. We must remember New York is a **populated** state (19 million people). At a 2% crime rate, that is nearly **half a million** crime cases annually!



**Using historical data, can  
we help the NYPD  
become more proactive  
and efficient?**

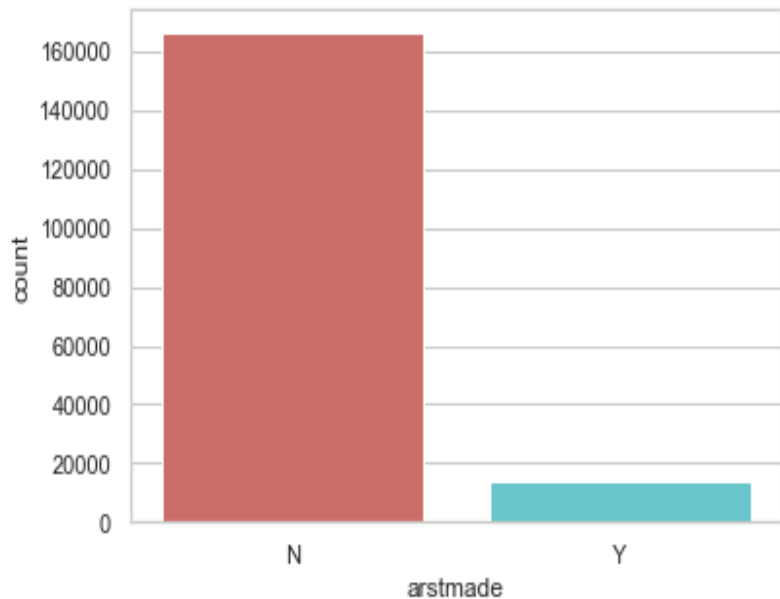
# Data utilized

- NYPD Stop, Question and Frisk Data (NYC.gov)

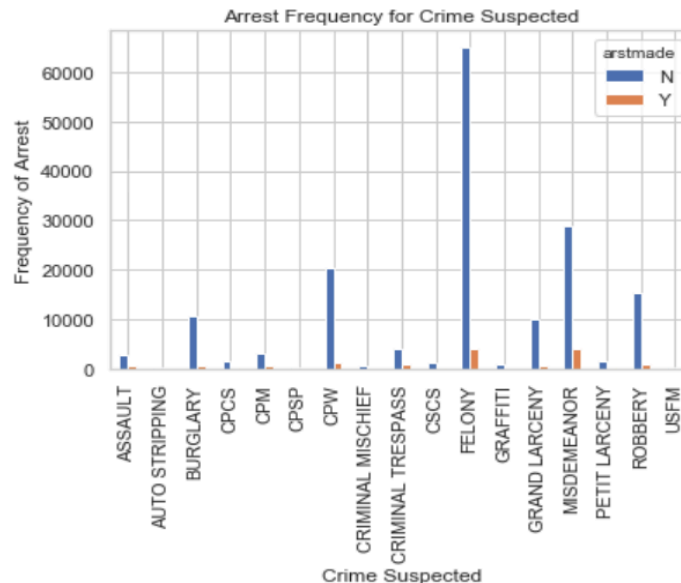


**NYPD**

# Exploratory Data Analysis

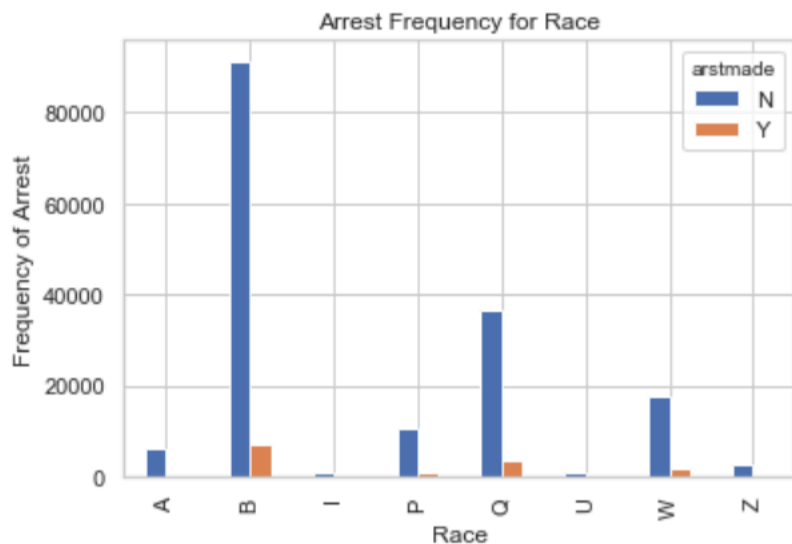


*NYPD **only** arrested 7% of people they stopped*

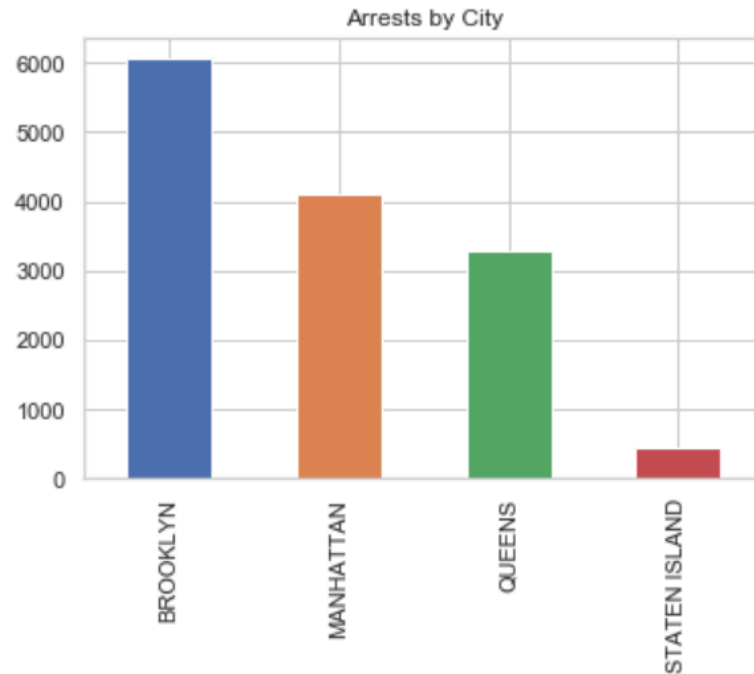


*Certain crimes did NOT lead to arrests*

# Exploratory Data Analysis (continued)



*Race does seem to play a role  
in arrest decisions*

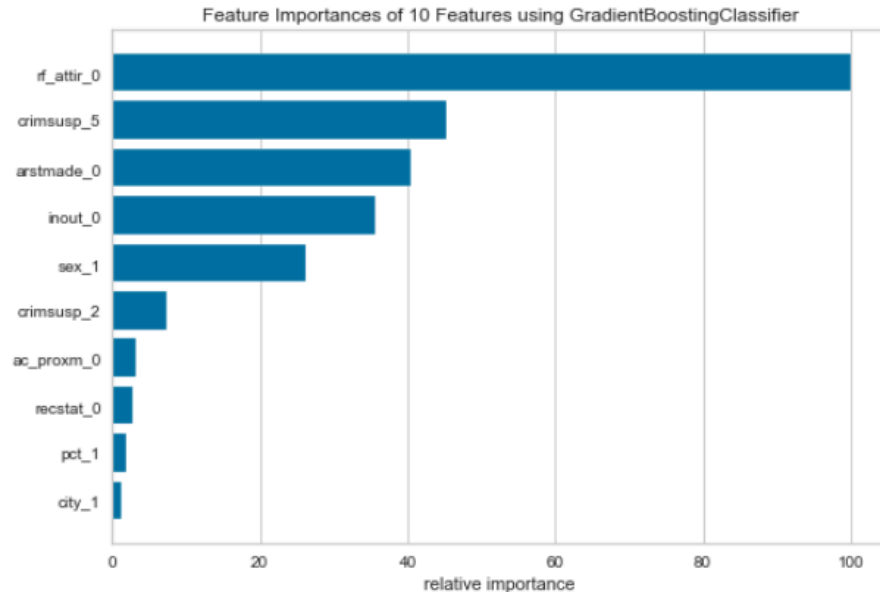


*Seems **Brooklyn** will require additional  
staffing versus **Staten Island***

# Modeling

MODEL	F-Measure
Linear Regression	.66
Random Forest	.67
K-Nearest Neighbors	.60

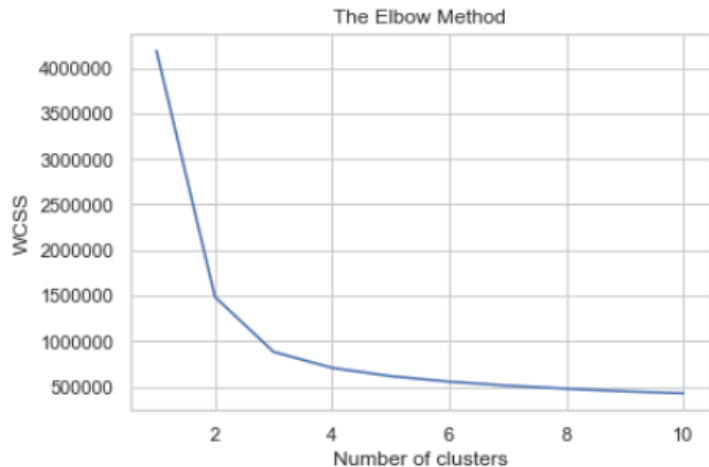
*Classification model performance*



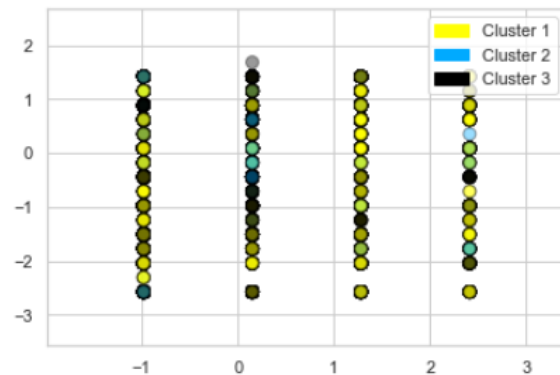
*Gradient Boosting to discover top 10 features in model*

# Clustering

Cluster analysis segments data into groups, so the NYPD understands what criminal “profiles” can look like.



*Within-Cluster Sum of Squares  
suggests we use 3 clusters*



**Cluster 1 – 59% of data**

- 29 y/o, male, Queens residents

**Cluster 2 – 8% of data**

- 29 y/o, male, Brooklyn & Manhattan residents

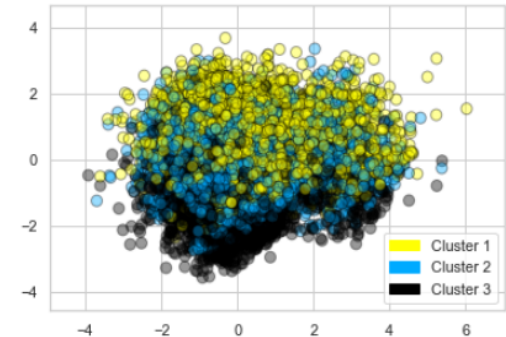
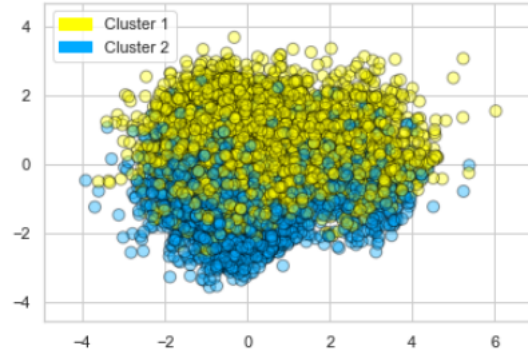
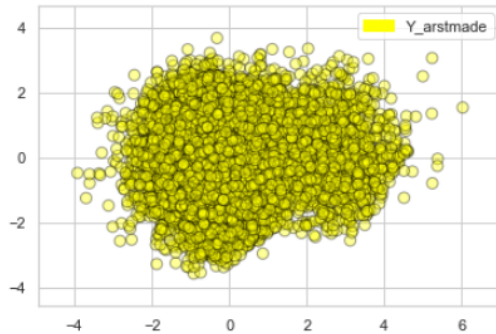
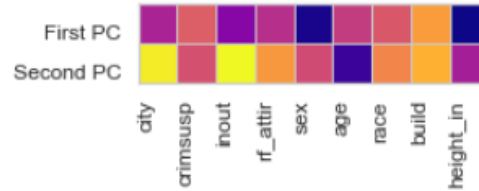
**Cluster 3 – 33% of data**

- 32 y/o, female, Manhattan



# Principal Component Analysis

I use dimension reduction technique, PCA, to summarize features in the data.



# FINAL CONCLUSIONS

Let's sum up the matter with some final conclusions about the presentation.

## Recommendation

Allocate more police force in Brooklyn and Queens. Focus on crimes with higher arrest rate (felony).

## Important Features

Attire and suspected crime are huge red flags to look out for in potential criminals

## Ethics

Information should remain private and kept for internal-use only.



**Thank you!**